Supplement to
The AMERICAN
PHYSICS TEACHER

VOLUME 2 SEPTEMBER, 1934 NUMBER 3, SUPPLEMENT

*The report of the Committee on Tests of the American Association of Physics Teachers which appears in this Supplement was not completed in time to be included in the September issue. Because many inquiries concerning the outcome of the testing program have been received and the report contains recommendations of importance for the successful continuation of the program, the Committee felt that the report should be published in advance of the December issue, preferably before the beginning of the new academic year. Its appearance in a Supplement to the September issue became possible, however, only after a successful appeal was made to the Cooperative Test Service and the Committee on Educational Testing of the American Council on Education for funds to defray the expense of publication.*

*To insure the future success of this testing program, the most elaborate ever undertaken in any college subject-matter field, it is essential that departments of physics continue their splendid cooperation and that individuals contribute by offering constructive criticisms of the present report. Comments addressed to this journal will be considered for inclusion in a symposium on the testing program planned for an early issue.*—THE EDITOR.

---

## The 1933–1934 College Physics Testing Program

### I. INTRODUCTION

IN the December, 1933, issue of *The American Physics Teacher*, President Karl T. Compton commented upon the proposed survey of college physics, in part, as follows:

> It is reasonable that this new venture is but the first of a series which will stimulate interest in and improve the quality of the teaching of college physics. . . . Being unofficial and optional, it does not curtail the freedom of any school to set its own standards and give its own examinations. . . . Speaking for the American Institute of Physics, I am sure that this new move will be generally approved and its broader educational results watched with interest.

The proposal itself, submitted by the Committee on Tests of the American Association of Physics Teachers, explained the purpose of the program:

> We recognize that any attempt to evaluate achievement is surrounded with dangers . . . but the attempt, if undertaken strictly from an experimental standpoint, is enticing. . . . We believe that the subject matter of physics itself and the traditional course examinations afford an excellent point of departure for the kind of survey here suggested and that the principle of measurement so fundamental to physics may be applied advantageously to the results of course instruction in the field. Recent nation-wide testing programs conducted under the auspices of the American Council on Education have disclosed the fact . . . that enormous variability obtains in all colleges on science, language and culture tests so that even the best colleges have some poor students and the poorest colleges have some very good students. . . . The inference is that similar differences may obtain in physics. If these differences are shown

to exist, many interesting and helpful suggestions may follow without threatening the art of teaching college physics with the deadly blight of standardization.

That the physicists engaged in college teaching entered whole-heartedly into this venture is testified by the fact that no less than 355 departments supported the program. The committee thanks these collaborators. Other indebtedness is acknowledged to two agencies of the American Council on Education; namely, the Cooperative Test Service and the Committee on Educational Testing. Thanks are also due to Miss Bette March for many tabulations and help in interpretation.

It should be noted in passing that not all of the 355 departments reported their results and that some made returns too late to be included in the table of norms. Differences in numbers in Table I and Table III are thus accounted for.

At the outset, the Committee on Tests of the American Association of Physics Teachers had as one of its objectives an inquiry into *whether what students know is identical with what they have been taught.* A host of letters and suggestions confirmed the wisdom of choosing such an objective and dozens of specific recommendations revealed the fact that physicists were interested in finding out whether prerequisites, number of hours spent in lecture and laboratory, textbooks used and, in general, formal educational machinery, contributed significantly to group differences in performance; or, group differences being small, whether individual students varied independently of formal requirements and classroom procedures in acquiring a knowledge of physics, at least as measured by the tests.

The committee thoroughly recognizes the limitations of the survey and also the fact that no more than hypothetical conclusions can be drawn. But as far as the data permit interpretation, two generalizations seem pertinent:

(1) that less confidence should be placed in prerequisites, formal classroom and laboratory procedures, textbooks and the like;

(2) that more effort should be directed to individual diagnosis and to providing opportunity for self-initiated and self-propelled study. For such work formal recognition might well be given.

The 355 colleges ordering tests are listed below by states:

*Alabama:* Athens College, Spring Hill College,* Talladega College.*

*Arizona:* University of Arizona, State Teachers College, Flagstaff.

*Arkansas:* Arkansas Polytechnic Institute,* Harding College, Hendrix College,* University of Arkansas.

*California:* Bakersfield Junior College,* California Institute of Technology,** Central Junior College, Citrus Junior College, Glendale Junior College,* Long Beach Junior College,* Loyola University of Los Angeles,* Menlo Junior College,* Occidental College,** Pacific Union College,* Pasadena Junior College, Riverside Junior College,* San Bernardino Junior College, South California Junior College, University of Redlands, Ventura Junior College.*

*Colorado:* Colorado College,* Colorado Vocational School,* State Junior College.

*Connecticut:* Connecticut State College, Trinity College, U. S. Coast Guard Academy, Wesleyan University.

*District of Columbia:* Wilson Teachers College.**

*Florida:* University of Miami.

*Georgia:* Agnes Scott College, Emory University, Emory Junior College, Oxford; Emory Junior College, Valdosta; Morehouse College,* University of Georgia.*

*Idaho:* College of Idaho,* University of Idaho.

*Illinois:* Armour Institute of Technology, Augustana College,* Aurora College, Broadview Junior College, College of St. Francis,* DePaul University, Eureka College, Illinois State Normal College, Illinois Wesleyan University, James Millikan University, Joliet Junior College, Knox College, Lake Forest College, LaSalle-Peru-Oglesby Junior College, Lincoln College, Loyola University,* Monmouth College,* Morton Junior College,** North Central College,* North Park College,* Northern Illinois State Teachers College,* Rockford College, Springfield Junior College, Thornton Junior College,* University of Chicago (Ryerson Laboratory),* University of Illinois, Western Illinois State Teachers College, Wheaton College.

*Indiana:* Central Normal College,* DePauw University, Earlham College, Franklin College, Goshen College, Hanover College,* Huntington College, Manchester College, Purdue University, St. Joseph's College, St. Mary-of-the-Woods College,* Taylor University,* Wabash College.*

*Iowa:* Central College,* Cornell College, Fletcher College, Fort Dodge Junior College, Graceland College, Grinnell College, Iowa State College, Iowa Wesleyan College, Maquoketa Junior College, Parsons College, St. Ambrose College, Simpson College, State University of Iowa, University of Dubuque, Upper Iowa University,* Waldorf Junior College, Wartburg College, Western Union College.*

*Kansas:* Baker University, Bethel College,* Central College, College of Emporia, Ft. Hays State Teachers College, Friends University, Garden City Junior College,* Hutchinson Junior College, Kansas State College,* Kansas Wesleyan College, McPherson College, Ottawa University, St. Benedict's College,* University of Wichita,* Washburn College.*

*Kentucky:* Berea College, Eastern Kentucky State Teachers College, Louisville Municipal College,* Lindsey Wilson Junior College, University of Louisville, Western Kentucky State Teachers College.*

*Louisiana:* Louisiana College,* Louisiana State University, Tulane University.

*Maine:* Bates College, Bowdoin College, Colby College, University of Maine.

*Maryland:* Goucher College, Hood College,* Johns Hopkins University,* Morgan College,* Washington College, Western Maryland College, Woodstock College,* Loyola College.

*Massachusetts:* Amherst College, Atlantic Union College,* International Y. M. C. A. College,* Lasell Junior College, Mount Holyoke College, Simmons College, Smith College, Wellesley College.*

*Michigan:* Albion College, Alma College, Battle Creek College, Bay City Junior College,** Ferris Institute,* Flint Junior College, Jackson College, Kalamazoo College, Michigan State Normal College, Olivet College, Port Huron Junior College, University of Michigan,** Western State Teachers College.

*Minnesota:* College of St. Benedict, College of St. Catherine, College of St. Thomas,** Concordia College, Duluth Junior College,** Ely Junior College,** Eveleth Junior College, Hibbing Junior College, Itasca Junior College, St. John's University, St. Paul Luther College, University of Minnesota,* Virginia Junior College, Winona State Teachers College.

*Mississippi:* University of Mississippi.*

*Missouri:* Culver Stockton College,** Maryville College,* Missouri Valley College, Moberly Junior College, Northeast Junior College, Kansas City; Park College, The Principia,* St. Joseph Junior College, State Teachers College,** Kirksville; University of Missouri, Washington University, William Jewell College.

*Montana:* Montana State College,* Carroll College, Northern Montana College.

*Nebraska:* Creighton University,* Dana College, Doane College,* Hebron College, Municipal University of Omaha, Nebraska Wesleyan University, State Teachers College, Kearney; Union College, York College.

*New Jersey:* Georgian Court College, N. J. State Teachers College, St. Peter's College, Upsala College.

*New Mexico:* New Mexico Military Institute, New Mexico Normal University, New Mexico State College.*

*New York:* Alfred University, Brooklyn College,** Colgate University, College of Mt. St. Vincent,* Columbia College, Cornell University,* Fordham College, Hamilton College, Hobart College, Houghton College, Hunter College,* Keuka College, Niagara University, St. Stephen's College, Skidmore College, Syracuse University, Union College,* Vassar College, Wells College.

*North Carolina:* Biltmore Junior College,* Catawba College, Duke University, Elon College,* Guilford College, North Carolina State College,* St. Augustine's College, University of North Carolina, Women's College of the University of North Carolina.

*North Dakota:* Jamestown College, North Dakota School of Forestry, State Teachers College,** Minot; University of North Dakota.

*Ohio:* Antioch College, Baldwin Wallace College, Bowling Green State College, Defiance College, Denison University, Heidelberg College, John Carroll University,* Kenyon College,* Marietta College, Miami University, Muskingum College, Notre Dame College, Ohio University, Ohio Wesleyan University,** Otterbein College,* St. John's College,* Wittenberg College, Xavier University.

*Oklahoma:* Bethany Peniel College, Cameron State Agricultural College, Oklahoma City University, Murray State School of Agriculture,* Oklahoma A. and M. College,* Okmulgee Junior College,** Panhandle A. and M. College, Southwestern State Teachers College,* University Preparatory School,* University of Oklahoma.

*Oregon:* Linfield College, University of Oregon.

*Pennsylvania:* Allegheny College, Bryn Mawr College, Clarion State Teachers College, College Misericordia,* Duquesne University, Geneva College, Gettysburg College,* Grove City College, Haverford College, Immaculata College,* Juniata College,* Lafayette College, LaSalle College, Lincoln University, Moravian College, Muhlenberg College, Pennsylvania State College, St. Joseph's College, St. Vincent College, Shippensburg State Teachers College, State Teachers College,* California; State Teachers College, Lock Haven; State Teachers College,* Mansfield; State Teachers College, Millersville; State Teachers College,* West Chester; Swarthmore College, Thiel College, University of Pennsylvania; University of Pittsburgh, Erie Center; University of Pittsburgh, Johnstown Center; University of Pittsburgh, Pittsburgh; Ursinus College,** Waynesburg College, Wilson College.

*Rhode Island:* Brown University, Providence College, Rhode Island State College.**

*South Carolina:* Clemson Agricultural College, Coker College,** Erskine College,* Presbyterian College, The Citadel.

*South Dakota:* Augustana College,* Northern State Teachers College.

*Tennessee:* Fisk University, Hiwassee College,* Nashville Agricultural Normal Institute,* Madison; Southwestern, Tusculum College, Union University,** University of the South.

*Texas:* A. and M. College of Texas, Amarillo Junior College,* Houston Junior College, John Tarleton Agric. College,* Paris Junior College, St. Edward's University,* Texas Christian University, Texas Lutheran College, Texas State College for Women,* Texas Technological College, University of Texas.*

*Utah:* Brigham Young University,** University of Utah, Utah State Agricultural College.

*Virginia:* Hampton Institute, Mary Baldwin College, University of Richmond,* University of Virginia.

*Washington:* College of Puget Sound, Gonzaga University, St. Martin's College, State College of Washington,* University of Washington, Walla Walla College, Yakima Valley Junior College.

*West Virginia:* Alderson-Broaddus College, Concord State Teachers College, West Liberty State Teachers College, West Virginia University.

*Wisconsin:* Lawrence College, Mission House College, St. Norbert College, State Teachers College, LaCrosse; State Teachers College,* Milwaukee; State Teachers College, River Falls; State Teachers College, Superior; University of Wisconsin.**

* Colleges not included in national distributions.
** Colleges not included in national distributions but included in distributions of averages and other parts of the report.

## II. The Tests

### Structure of tests

The tests were constructed to sample the content of elementary physics and to permit the making of comparable forms. Thus, one form could be used before any given course in elementary physics, and another form could be used subsequently to measure growth in the subject. The difference obtained might then be regarded as a tentative norm of expectancy for future classes. Originally mechanics, heat and sound were planned for one combined form, and light, electricity and modern physics for the other, since these combinations represented the normal progression in most courses. But subsequently, in answer to demands for atypical combinations, the six subjects were printed separately as follows: mechanics, 48 items, 60 minutes; heat, 27 items, 30 minutes; sound, 16 items, 20 minutes; light, 35 items, 40 minutes; electricity, 44 items, 50 minutes; modern physics, 19 items, 25 minutes. (Hereafter in the report these subjects will be abbreviated thus: M, H, S, L, E, Mp.) The committee gladly acceded to the wishes of the departments for separately printed topics but it is suggested that specialization in physics usually follows and is predicated upon the elementary course as a whole. Since the study reported below reveals large individual differences regardless of formal training, it is recommended

Table I. *Difficulty and validity indices for each item in the M, H and S, 1933 test forms and L, E and Mp, 1934 test forms.**

| Item no. | *Mechanics* Difficulty (percent correct) | Validity index | Item no. | *Heat* Difficulty (percent correct) | Validity index | Item no. | *Sound* Difficulty (percent correct) | Validity index |
|---|---|---|---|---|---|---|---|---|
| 1 | 90 | 3 | 1 | 74 | 6 | 1 | 83 | 2 |
| 2 | 63 | 5 | 2 | 68 | 6 | 2 | 64 | 5 |
| 3 | 91 | 2 | 3 | 62 | 6 | 3 | 62 | 6 |
| 4 | 85 | 2 | 4 | 74 | 6 | 4 | 75 | 4 |
| 5 | 61 | 2 | 5 | 65 | 3 | 5 | 65 | 3 |
| 6 | 75 | 5 | 6 | 75 | 4 | 6 | 77 | 4 |
| 7 | 69 | 2 | 7 | 69 | 4 | 7 | 51 | 5 |
| 8 | 49 | 2 | 8 | 65 | 3 | 8 | 67 | 4 |
| 9 | 71 | 5 | 9 | 58 | 6 | 9 | 41 | 6 |
| 10 | 64 | 3 | 10 | 66 | 6 | 10 | 58 | 4 |
| 11 | 46 | 5 | 11 | 47 | 9 | 11 | 58 | 4 |
| 12 | 71 | 4 | 12 | 48 | 2 | 12 | 45 | 5 |
| 13 | 69 | 5 | 13 | 52 | 3 | 13 | 60 | 5 |
| 14 | 66 | 7 | 14 | 40 | 7 | 14 | 33 | 3 |
| 15 | 75 | 5 | 15 | 39 | 4 | 15 | 39 | 7 |
| 16 | 62 | 2 | 16 | 60 | 4 | 16 | 41 | 4 |
| 17 | 71 | 4 | 17 | 48 | 4 | | | |
| 18 | 45 | 5 | 18 | 20 | 3 | | | |
| 19 | 55 | 5 | 19 | 35 | 3 | | | |
| 20 | 63 | 5 | 20 | 41 | 6 | | | |
| 21 | 74 | 3 | 21 | 36 | 5 | | | |
| 22 | 53 | 5 | 22 | 38 | 4 | | | |
| 23 | 41 | 4 | 23 | 30 | 5 | | | |
| 24 | 48 | 4 | 24 | 41 | 4 | | | |
| 25 | 61 | 5 | 25 | 24 | 4 | | | |
| 26 | 79 | 5 | 26 | 37 | 5 | | | |
| 27 | 40 | 7 | 27 | 22 | 5 | | | |
| 28 | 41 | 3 | | | | | | |
| 29 | 45 | 5 | | | | | | |
| 30 | 48 | 4 | | | | | | |
| 31 | 70 | 5 | | | | | | |
| 32 | 58 | 4 | | | | | | |
| 33 | 54 | 7 | | | | | | |
| 34 | 49 | 6 | | | | | | |
| 35 | 31 | 2 | | | | | | |
| 36 | 33 | 8 | | | | | | |
| 37 | 33 | 4 | | | | | | |
| 38 | 45 | 5 | | | | | | |
| 39 | 25 | 4 | | | | | | |
| 40 | 24 | 5 | | | | | | |
| 41 | 12 | 3 | | | | | | |
| 42 | 24 | 4 | | | | | | |
| 43 | 6 | −2 | | | | | | |
| 44 | 30 | 4 | | | | | | |
| 45 | 19 | 5 | | | | | | |
| 46 | 8 | 1 | | | | | | |
| 47 | 15 | 4 | | | | | | |
| 48 | 9 | 3 | | | | | | |

that departments continue combinations of the tests for diagnostic purposes and for identifying students whose general grasp of elementary physics is unusual. This practice seems more reasonable for the selection of advanced students and even for identifying special interests than one which would restrict measurement too narrowly to those topics of peculiar interest to a given instructor.

One form each of M, H and S was provided at the end of the first semester, 1933 forms. Comparable 1934 forms were made available later in the academic year. At the beginning of the second semester, 1933 forms of L, E and Mp were printed and comparable 1934 forms were subsequently issued for end-semester testing. The two general objectives in view were to conduct a survey of elementary physics and to measure student gains over the period of at least one semester.

**Difficulty and validity of individual test questions**

That the tests were satisfactorily scaled for difficulty and for validity of items appears from inspection of Table I. Ideally tests should include items representing all levels of difficulty, from very easy to very hard, as determined by percentage of right answers, and each item included should be at least reasonably valid; that is, it should differentiate between students making a high total score and those making a low total score. How nearly the tests approximate this ideal can be judged from Tables I and II. The numerical values for the validity indices represent quarter-sigma units. (For an explanation of the validity index see paragraph 6 of the next section.)

TABLE I. (*Continued.*)

| *Light* | | | *Electricity* | | | *Modern Physics* | | |
|---|---|---|---|---|---|---|---|---|
| Item no. | Difficulty (percent correct) | Validity index | Item no. | Difficulty (percent correct) | Validity index | Item no. | Difficulty (percent correct) | Validity index |
| 1 | 79 | 8 | 1 | 89 | 4 | 1 | 51 | 2 |
| 2 | 79 | 8 | 2 | 85 | 4 | 2 | 84 | 3 |
| 3 | 88 | 3 | 3 | 74 | 9 | 3 | 75 | 5 |
| 4 | 71 | 6 | 4 | 70 | 5 | 4 | 64 | 3 |
| 5 | 61 | 8 | 5 | 68 | 5 | 5 | 68 | 4 |
| 6 | 60 | 10 | 6 | 62 | 5 | 6 | 45 | 3 |
| 7 | 57 | 6 | 7 | 48 | 7 | 7 | 58 | 4 |
| 8 | 68 | 4 | 8 | 62 | 5 | 8 | 45 | 5 |
| 9 | 64 | 6 | 9 | 58 | 6 | 9 | 38 | 5 |
| 10 | 46 | 5 | 10 | 65 | 6 | 10 | 31 | 5 |
| 11 | 43 | 5 | 11 | 52 | 3 | 11 | 43 | 1 |
| 12 | 35 | 3 | 12 | 64 | 5 | 12 | 31 | 4 |
| 13 | 47 | 6 | 13 | 57 | 6 | 13 | 18 | 2 |
| 14 | 47 | 4 | 14 | 64 | 5 | 14 | 19 | 1 |
| 15 | 46 | 7 | 15 | 55 | 3 | 15 | 28 | 6 |
| 16 | 38 | 4 | 16 | 53 | 5 | 16 | 20 | 3 |
| 17 | 43 | 6 | 17 | 68 | 5 | 17 | 12 | 4 |
| 18 | 45 | 5 | 18 | 56 | 5 | 18 | 30 | 4 |
| 19 | 31 | 5 | 19 | 56 | 8 | 19 | 42 | 5 |
| 20 | 34 | 6 | 20 | 58 | 6 | | | |
| 21 | 36 | 4 | 21 | 38 | 3 | | | |
| 22 | 36 | 5 | 22 | 40 | 4 | | | |
| 23 | 40 | 5 | 23 | 42 | 4 | | | |
| 24 | 36 | 5 | 24 | 54 | 5 | | | |
| 25 | 32 | 4 | 25 | 53 | 7 | | | |
| 26 | 26 | 4 | 26 | 44 | 2 | | | |
| 27 | 27 | 6 | 27 | 43 | 5 | | | |
| 28 | 29 | 5 | 28 | 35 | 5 | | | |
| 29 | 23 | 6 | 29 | 33 | 6 | | | |
| 30 | 19 | 4 | 30 | 36 | 2 | | | |
| 31 | 29 | 3 | 31 | 32 | 4 | | | |
| 32 | 17 | 4 | 32 | 36 | 5 | | | |
| 33 | 40 | 5 | 33 | 30 | 5 | | | |
| 34 | 16 | 5 | 34 | 61 | 5 | | | |
| 35 | 14 | 2 | 35 | 36 | 5 | | | |
| | | | 36 | 33 | 6 | | | |
| | | | 37 | 19 | 8 | | | |
| | | | 38 | 12 | 3 | | | |
| | | | 39 | 21 | 5 | | | |
| | | | 40 | 18 | 5 | | | |
| | | | 41 | 19 | 4 | | | |
| | | | 42 | 11 | 2 | | | |
| | | | 43 | 13 | 7 | | | |
| | | | 44 | 18 | 4 | | | |

* The test papers analyzed were those of 2000 students in each semester who had taken all three tests. The difficulty ratings are percentages of students who answered each item correctly. The validity index indicates the degree to which an item discriminates between the high- and low-scoring students. The groups of high and low students were selected on the basis of total scores for the tests of each semester. A validity index of 0 means that as many poor students answered the item correctly as good students; a minus sign indicates that poor students answered correctly more often than good students. Indices of 2 indicate satisfactory items; indices of 4 or above are very good items.

TABLE II. *Summary from Table I of difficulty and validity indices.*

| Difficulty | M | H | S | L | E | Mp |
|---|---|---|---|---|---|---|
| 90 | 2 | – | – | – | – | – |
| 85 | 1 | – | – | 1 | 2 | – |
| 80 | – | – | 1 | – | – | 1 |
| 75 | 3 | 1 | 2 | 2 | – | 1 |
| 70 | 5 | 2 | – | 1 | 2 | – |
| 65 | 3 | 5 | 2 | 1 | 3 | 1 |
| 60 | 6 | 2 | 3 | 3 | 5 | 1 |
| 55 | 2 | 1 | 2 | 1 | 6 | 1 |
| 50 | 2 | 1 | 1 | – | 4 | 1 |
| 45 | 8 | 3 | 1 | 5 | 1 | 2 |
| 40 | 3 | 3 | 2 | 4 | 4 | 2 |
| 35 | – | 5 | 1 | 5 | 6 | 1 |
| 30 | 4 | 1 | 1 | 3 | 3 | 3 |
| 25 | 1 | – | – | 4 | – | 1 |
| 20 | 2 | 3 | – | 1 | 1 | 1 |
| 15 | 2 | – | – | 3 | 4 | 2 |
| 10 | 1 | – | – | 1 | 3 | 1 |
| 5 | 3 | – | – | – | – | – |
| **Validity Index** | **M** | **H** | **S** | **L** | **E** | **Mp** |
| 10 | – | – | – | 1 | – | – |
| 9 | – | 1 | – | – | 1 | – |
| 8 | 1 | – | – | 3 | 2 | – |
| 7 | 3 | 1 | 1 | 1 | 3 | – |
| 6 | 1 | 7 | 2 | 8 | 6 | 1 |
| 5 | 17 | 4 | 4 | 10 | 18 | 5 |
| 4 | 11 | 8 | 6 | 8 | 7 | 5 |
| 3 | 6 | 5 | 2 | 3 | 4 | 4 |
| 2 | 7 | 1 | 1 | 1 | 3 | 2 |
| 1 | 1 | – | – | – | – | 2 |
| 0 | – | – | – | – | – | – |
| −1 | – | – | – | – | – | – |
| −2 | 1 | – | – | – | – | – |

## Statistical terms

Any adequate interpretation of the data in this report depends, in part, upon an understanding of certain statistical concepts. Statistics is an applied science and, because its vocabulary is specific, it seems advisable to include definitions of terms and expressions which may otherwise suffer from uncertainty. Many readers of this report are sufficiently familiar with statistical methods and terminology not to need a glossary; they may pass over the following six paragraphs,[1] although the description of item analysis (6) may be of general interest.

1. The *coefficient of correlation*, $r$, is a mathematical expression of the degree of association or of interdependence between paired variables. The equation is set up so that a quantity may be derived which varies from a minimum value of 0 to a maximum value of $\pm 1$, as the association varies from nothing to perfect rectilinear interdependence. A perfect negative, or a perfect positive correlation is thus expressed by $-1$ or $+1$, respectively, and complete absence of correlation by 0. By this means, the relation between the paired variates of two sets of variables may be determined precisely. In the interpretation of correlation coefficients, it must be kept in mind that increments in the magnitude of the correlation coefficient are not to be regarded as arithmetical percentages of increase in relationship between the two variables. That is, increase in the magnitude of $r$ is not indicative of a point for point increase in the degree of relationship. The increase in an $r$ from .20 to .30 is relatively much smaller than an increase from .70 to .75. For practical purposes, correlations in the range $\pm .30$ may be said to indicate negligible association between variables. Only when the correlation, whether positive or negative, is equal to .60 or more, can much practical significance be attached to the figures. The equation for the correlation may be expressed:

$$r = ((\Sigma XY/N) - M_x M_y)/\sigma_x \sigma_y$$

where $\Sigma XY$ is the sum of the products of the two variables $X$ and $Y$; $M_x$ and $M_y$ are the arithmetic means of the variable; and $\sigma_x$ and $\sigma_y$ are the standard deviation of each.

2. *Percentile* scaling is a method of scaling raw scores to make their meaning more immediately apparent. For example, if 10,000 students have taken a test, each raw score is given a value on a scale of 100 points. All the scores of the 10,000 students are arranged in order of magnitude and the lowest 1 percent of these scores is assigned a percentile rank of 1, the next 1 percent a percentile rank of 2, etc. Accordingly, if a student has a percentile rank of 60, this means that of 10,000 students tested with him on the same examination 60 percent received scores equal to, or lower than, his score and 40 percent received higher scores. The percentile scale derives its chief value from the fact that it expresses test scores on a comparable, relative scale. That is, a student receives raw scores of 80 on two tests and thus appears equally high on both; but when converted into percentile ranks, the score of 80 may mean that on one test his achievement was better than that of 90 percent of other students tested with him, and on the other better than that of only 60 percent.

3. The *median* is the central value of a series of scores when the scores are arranged in order of numerical magnitude. It is a measure of central tendency or an average that is little affected by extreme scores.

4. The *mean*, $M$, is the arithmetic average. Since it is a function of the magnitude of the scores and the number of the scores, it is sensitive to the extreme (very high and very low) scores in a distribution.

5. The *standard deviation*, or *sigma* ($\sigma$), is the root-mean-square deviation. It may conveniently be considered as a value on the base-line of a univariate Gaussian distribution, between which and the mean lie approximately 34 percent of the cases. Between the points 1 $\sigma$ below the mean and 1 $\sigma$ above the mean are included 68 percent of the cases.

[1] Detailed definitions of statistical terms may be found in K. J. Holzinger, *Statistical Methods for Students in Education*, Ginn, 1928; T. L. Kelley, *Statistical Method*, Macmillan, 1924; A. E. Treloar, *Outlines of Biometric Analysis*, Burgess-Rosebury Co., 1933; and G. U. Yule, *An Introduction to the Theory of Statistics*, Charles Griffin & Co., 1929.

The range thus measured indicates the extent to which scores tend to cluster around the mean or to spread away from it. The working formula may be expressed as follows: $((\Sigma X^2/N)-M_x^2)^{\frac{1}{2}}$ where $\Sigma X^2$ is the sum of the squared scores, $N$ is the number of the scores and $M_x^2$ is the squared mean of the scores.

6. *Validity indices* of individual test items were derived by the following procedure. It is assumed that those students who get the highest scores on a carefully made test have a better understanding of the subject, as measured by the tests, than those who get low scores. Hence, for each item the percentage of right answers of the high-scoring group is plotted against the percentage of right answers for the low-scoring group as in Fig. 1.

Inspection of Fig. 1 reveals that item number 1 differentiates well between the high-scoring and the low-scoring groups, since 98 percent of the former and only 2 percent of the latter answer it correctly. Item 2 scarcely differentiates between these groups at all. Item 3 reverses the normal expectancy; that is, more poor students answer it correctly than do good students. Since only items which have good differentiating power should be retained in a general achievement test, items 2 and 3 should be eliminated. Item 4 not only fails to differentiate but is very easy, perhaps too easy. Item 2 fails to differentiate but is of medium difficulty; and item 5 not only fails to differentiate but is extremely difficult. In the item analysis of the physics tests, the diagonal which bisects the graph was given the value of zero and the degree of deviation of each item from that diagonal determined its validity index. Therefore, a validity index (see Table I) of 0 means that as many poor students answer the item correctly as do good students; a minus sign indicates that poor students answered the item correctly more often than good students. Indices of 2 are satisfactory, while those of 4 or above have very good differentiating power. The percentages in Fig. 1 have been grouped to represent quarter-sigmas of a Gaussian distribution, in order to make more nearly comparable the percentage differences for items of all degrees of difficulty. The validity indices of Tables I and II are, therefore, differences expressed in terms of quarter-sigmas.

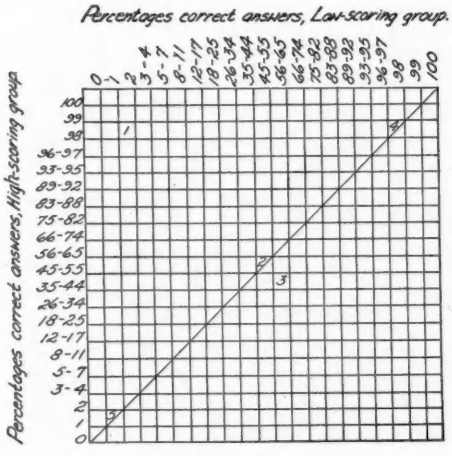


FIG. 1. Method of item analysis.

## Validity and reliability of tests

According to the judgment of the great majority of the physics teachers participating in this experiment, the tests are valid in the sense that they measure an important aspect of achievement in elementary physics. This judgment is confirmed by correlations between final grades and test scores which have been calculated from data supplied by five colleges. The *r's* range from .66 to .88 as shown by the following summary:

| College | $r$ | N | *Mean* College grade | *Mean* M+H+S score | *Sigma* College grade | *Sigma* M+H+S score | Used test in final grade |
|---|---|---|---|---|---|---|---|
| 1 | .88 | 57 | 5.1 | 45.7 | 2.2 | 14.2 | No |
| 2 | .71 | 156 | 4.5 | 28.6 | 2.2 | 11.7 | Yes |
| 3 | .72 | 246 | 5.0 | 39.6 | 2.3 | 14.2 | Yes |
| | | | | L+E+Mp score | | L+E+Mp score | |
| 1 | .84 | 53 | 6.1 | 43.1 | 1.8 | 11.9 | No |
| 4 | .66 | 22 | 6.0 | 31.2 | 2.1 | 9.8 | No |
| 5 | .71 | 35 | 5.1 | 56.9 | 2.3 | 16.1 | Yes |

These *r's* between test scores and final semester grades are so high as to suggest that the tests might have been used as a partial basis for determining the final grades. Inquiry revealed the fact that three of the five colleges did use the tests for this purpose but the college that shows the highest *r's* for both semesters reported that the tests did not influence the final grades. The lowest *r* is .66 for College 4, in which the tests did not influence the grades. It seems obvious that these *r's* are high enough to indicate that the tests measure a function that physics teachers consider important in their final grades but they are not high enough to indicate mere duplication of grades.

TABLE III. *National percentiles, M, H, S, L, E and Mp.**

| | M | H | S | L | E | Mp | M+H | M+H+S | L+E | L+E+Mp | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No. Colleges | 180 | 169 | 93 | 112 | 108 | 61 | 85 | 76 | 97 | 54 | |
| No. Cases | 7957 | 7440 | 3741 | 4936 | 5149 | 1837 | 3598 | 3234 | 4305 | 1714 | |
| Mean | 18.3 | 11.0 | 8.1 | 12.5 | 17.3 | 6.2 | 28.7 | 38.2 | 29.9 | 37.3 | |
| Sigma | 8.8 | 5.6 | 3.9 | 5.7 | 6.9 | 3.7 | 13.0 | 16.3 | 11.4 | 14.8 | |
| Percentile | | | | | | | | | | | Percentile |
| 100 | 48 | 27 | 16 | 33 | 43 | 19 | 71 | 86 | 73 | 91 | 100 |
| 99 | 40 | 24 | (16) | 27 | 35 | 16 | 61 | 76 | 59 | 76 | 99 |
| 98 | 38 | 23 | 15 | 25 | 33 | 15 | 58 | 73 | 56 | 71 | 98 |
| 97 | 36 | 22 | (15) | 24 | 31 | 14 | 55 | 71 | 53 | 69 | 97 |
| 96 | 35 | 21 | (15) | (24) | 30 | 13 | 53 | 68 | 51 | 66 | 96 |
| 95 | 34 | (21) | (15) | 23 | 29 | (13) | 52 | 66 | 50 | 64 | 95 |
| 94 | 33 | 20 | (15) | 22 | (29) | (13) | 51 | 65 | 49 | 62 | 94 |
| 93 | 32 | (20) | 14 | (22) | 28 | 12 | 49 | 63 | 48 | 61 | 93 |
| 92 | 31 | 19 | 13 | 21 | 27 | (12) | 48 | 62 | 47 | 60 | 92 |
| 91 | (31) | (19) | (13) | (21) | (27) | (12) | 47 | 61 | 46 | 58 | 91 |
| 90 | 30 | (19) | (13) | 20 | (27) | 11 | (47) | 60 | 45 | 57 | 90 |
| 88 | 29 | 18 | (13) | (20) | 26 | (11) | 45 | 58 | 43 | 55 | 88 |
| 86 | 28 | (18) | 12 | 19 | 25 | 10 | 44 | 56 | 42 | 53 | 86 |
| 84 | 27 | 17 | (12) | 18 | 24 | (10) | 42 | 55 | 41 | 52 | 84 |
| 82 | 26 | (17) | (12) | (18) | (24) | (10) | 41 | 53 | 40 | 51 | 82 |
| 80 | (26) | 16 | (12) | 17 | 23 | 9 | 40 | 52 | 39 | 49 | 80 |
| 75 | 24 | 15 | 11 | 16 | 22 | (9) | 37 | 49 | 37 | 47 | 75 |
| 70 | 23 | 14 | (11) | 15 | 21 | 8 | 35 | 47 | 35 | 44 | 70 |
| 65 | 21 | 13 | 10 | 14 | 20 | 7 | 33 | 44 | 34 | 42 | 65 |
| 60 | 20 | 12 | 9 | (14) | 19 | (7) | 31 | 42 | 32 | 40 | 60 |
| 55 | 19 | 11 | (9) | 13 | 18 | 6 | 29 | 40 | 31 | 38 | 55 |
| 50 | 18 | (11) | 8 | 12 | 17 | (6) | 28 | 38 | 29 | 36 | 50 |
| 45 | 17 | 10 | (8) | 11 | 16 | 5 | 26 | 36 | 28 | 34 | 45 |
| 40 | 16 | 9 | 7 | (11) | 15 | (5) | 24 | 34 | 26 | 32 | 40 |
| 35 | 14 | 8 | (7) | 10 | 14 | 4 | 23 | 31 | 25 | 30 | 35 |
| 30 | 13 | (8) | 6 | 9 | 13 | (4) | 21 | 28 | 24 | 28 | 30 |
| 25 | 12 | 7 | 5 | 8 | 12 | 3 | 19 | 26 | 22 | 26 | 25 |
| 20 | 11 | 6 | 4 | (8) | 11 | (3) | 17 | 24 | 20 | 24 | 20 |
| 18 | 10 | 5 | (4) | 7 | (11) | (3) | (17) | 22 | 19 | (24) | 18 |
| 16 | 9 | (5) | (4) | (7) | 10 | 2 | 16 | 21 | 18 | 23 | 16 |
| 14 | (9) | (5) | (4) | 6 | (10) | (2) | 15 | 20 | (18) | 22 | 14 |
| 12 | 8 | 4 | 3 | (6) | 9 | (2) | 14 | 19 | 17 | 20 | 12 |
| 10 | 7 | (4) | (3) | 5 | (9) | (2) | 12 | 17 | 16 | 19 | 10 |
| 9 | (7) | 3 | (3) | (5) | 8 | 1 | (12) | (17) | 15 | (19) | 9 |
| 8 | 6 | (3) | 2 | (5) | (8) | (1) | 11 | 16 | (15) | 18 | 8 |
| 7 | (6) | (3) | (2) | (5) | 7 | (1) | 10 | 15 | 14 | (18) | 7 |
| 6 | 5 | 2 | (2) | 4 | (7) | (1) | (10) | 13 | 13 | 17 | 6 |
| 5 | (5) | (2) | 1 | (4) | (7) | 0 | 9 | 12 | 12 | 16 | 5 |
| 4 | 4 | (2) | (1) | 3 | 6 | (0) | 8 | 11 | 11 | 14 | 4 |
| 3 | 3 | 1 | (1) | (3) | 5 | (0) | 7 | 9 | 10 | 13 | 3 |
| 2 | 2 | 0 | 0 | 2 | 4 | (0) | 6 | 8 | 9 | 11 | 2 |
| 1 | 0 | (0) | (0) | 1 | 3 | (0) | 4 | 6 | 7 | 9 | 1 |

* The scales are based upon returns for students tested after studying the various topics and show true percentiles, calculated from the distributions of scores available at the time of computation. Each score in each column shows the upper score limit of the percentile indicated at the extreme right and left of the line. For example, the bottom entry in the column for total score on M and H shows that all scores of 4 or below have a percentile value of 1; all scores of 5 or 6 have a percentile value of 2; and all scores above 61 have a percentile value of 100. Since colleges used varying combinations of tests, the numbers of cases and of colleges vary from column to column. The mean and sigma of the scores and the number of colleges involved are shown at the top of each column. When a score appears on a scale more than once, use the figure not in parenthesis.

The reliability coefficient[2] for the combined score of M, H and S is .91 and for L, E and Mp is .92, the number of cases being 300 for both.[3] The coefficients themselves indicate that the tests afford measures which are highly consistent. In other words, if two sets of these tests were given to the same class, the pupils would be put in substantially the same rank-order by one set of tests as by the other. The reliability coefficients of old-type essay tests of two or three hours length are usually lower than the foregoing ones found for the Cooperative Physics Tests. Many of the physics teachers in the experiment have suggested that the tests be lengthened; if this suggestion is carried out, both validity and reliability indices may be raised somewhat above their present satisfactorily high levels.

[2] Correlation between scores on odd against even numbered items to measure the consistency with which the items sample the content of elementary physics.

[3] The means and sigmas of the total scores of the samplings used in calculating these Spearman-Brown reliability coefficients are for M+H+S, 38.3 and 16.9 and for L+E+Mp, 39.6 and 17.5, respectively.

## III. THE TEST RESULTS

### National percentiles

Table III enables each college to find the percentile rank of each student on each topic. From this table it is evident that the tests are

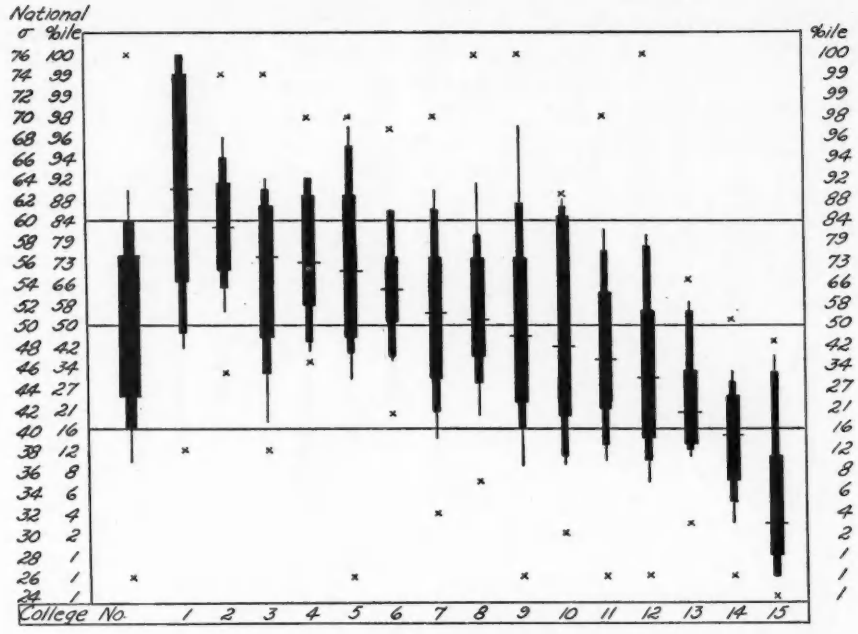


FIG. 2. Variability of achievement as measured by combined scores on M, H and S. The middle horizontal line shows the national median and the other two are at the 16th and 84th percentiles of the national distribution. The first vertical bar represents the combined national group and each of the other bars represents an individual college. The wide portion of each bar represents the range of scores of the middle half in each college. The narrow parts extend to the 16th and 84th percentile in each college, i.e., one standard deviation above and one below the mean. The lines at the ends extend down to the 10th percentile and up to the 90th percentile. The crosses below the bars represent the lowest scores and those above represent the highest scores in the several colleges (the range). The short cross line at the middle of each bar represents the median score of the college. Although this chart is based entirely on percentiles, the scale has been altered to correspond roughly to a sigma scale, so that vertical distances are approximately comparable. The sigma scale is derived from the percentile scale.

well adjusted in difficulty for participating colleges. The scores are not piled up at either end; hence, there is room at the top and bottom for differentiating between good and poor students.

**Variability of achievement**

The fifteen colleges in Fig. 2 were chosen to represent the whole range of medians from highest to lowest and all types of institutions reporting results in time for inclusion in the report. The fifteen institutions include engineering schools, universities, four-year colleges, junior colleges and teachers' colleges. Fig. 2 shows two types of variability; the first is the variability of median scores of students in individual colleges. The differences in student performance are very large indeed. So far as the functions measured by the combined M, H and S tests are concerned, the lowest group of colleges has almost nothing in common with the three or four colleges at the high end of the scale. However, the variability within colleges suggests that adjustment of course and pace to the needs and abilities of individual students must be made within the college, despite mitigation of the problem by progressive pre-college selection and guidance of students.

Fig. 3 illustrates for combined scores on L, E and Mp what Fig. 2 shows for M, H and S. Both are to be interpreted in the same way. It is obvious from both of these figures that the students in some colleges learn much less physics of the type measured by these tests than those in other colleges. In college 15, for example, all of the students are below the national average,

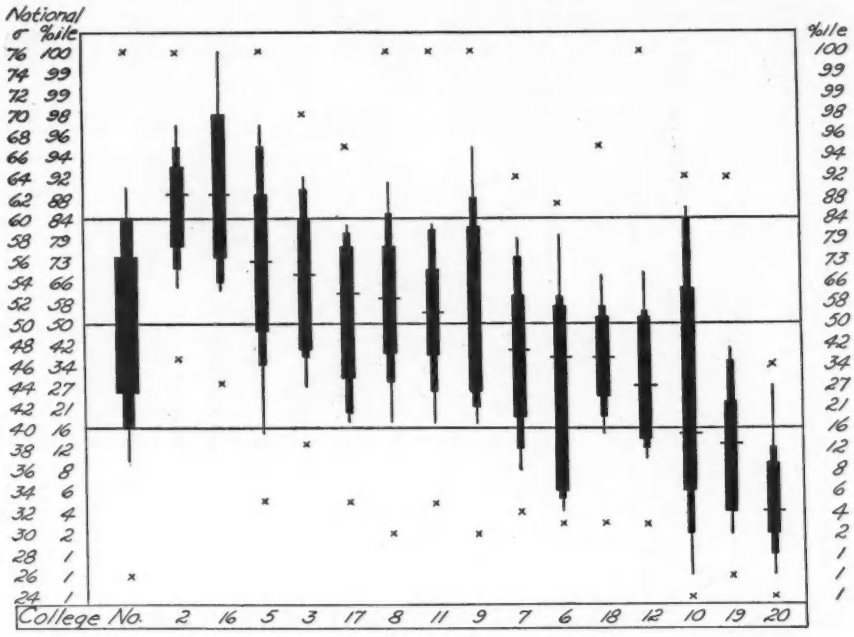


FIG. 3. Variability of achievement as measured by combined scores on L, E and Mp. To be read in the same manner as Fig. 2.

TABLE IV. *Distribution of college averages.**

| | M | H | S | L | E | Mp | M+H | M+H +S | L+E | L+E +Mp | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| National Percentiles | | | | | | | | | | | Percentiles |
| 96–97 | 1 | | | | | | | | 1 | | 96–97 |
| 94–95 | | | | 1 | 2 | | 1 | | | | 94–95 |
| 92–93 | 1 | 3 | 1 | 1 | 1 | | 1 | | 2 | | 92–93 |
| 88–91 | 5 | 3 | | 1 | 3 | 1 | 4 | 1 | 2 | 3 | 88–91 |
| 84–87 | 1 | 4 | 8 | 3 | 2 | 3 | 1 | 2 | 1 | | 84–87 |
| 79–83 | 8 | 9 | 10 | 6 | 2 | 3 | 3 | 3 | 4 | 1 | 79–83 |
| 73–78 | 20 | 20 | 11 | 5 | 7 | 9 | 7 | 11 | 3 | 4 | 73–78 |
| 66–72 | 17 | 31 | 26 | 15 | 21 | | 10 | 5 | 17 | 6 | 66–72 |
| 58–65 | 35 | 31 | 27 | 12 | 15 | 11 | 17 | 22 | 19 | 5 | 58–65 |
| 50–57 | 41 | 27 | 34 | 30 | 19 | 15 | 9 | 15 | 8 | 12 | 50–57 |
| 42–49 | 27 | 21 | 16 | 17 | 14 | 10 | 15 | 12 | 12 | 6 | 42–49 |
| 34–41 | 28 | 39 | 9 | 10 | 17 | 9 | 14 | 9 | 10 | 10 | 34–41 |
| 27–33 | 4 | 9 | 9 | 12 | 10 | | 3 | 6 | 11 | 4 | 27–33 |
| 21–26 | 12 | 15 | 1 | 5 | 4 | 1 | 7 | 1 | 5 | 3 | 21–26 |
| 16–20 | 10 | 2 | 2 | 4 | 6 | | 2 | 4 | 5 | 1 | 16–20 |
| 12–15 | 3 | 1 | 1 | 5 | 4 | 1 | 2 | 3 | 6 | 1 | 12–15 |
| 8–11 | 1 | 2 | | 2 | 3 | 1 | 1 | | | | 8–11 |
| 6–7 | | | | | | | | 1 | 1 | | 6–7 |
| 4–5 | | | | 1 | | | | | | 1 | 4–5 |
| No. Colleges | 214 | 217 | 155 | 130 | 130 | 64 | 97 | 95 | 107 | 57 | |
| No. Students | 8911 | 9104 | 6215 | 5673 | 6010 | 1887 | 4022 | 3675 | 4642 | 1762 | |

* The means on all six subjects and on four combinations are here distributed in terms of national percentiles. Although this distribution is based on percentiles, the scale has been altered so that the intervals correspond approximately to a sigma scale. The vertical distances are therefore roughly comparable. The numbers of colleges represented in this table are larger than in Table III, since all averages received on or before July 10, 1934, are included.

while in college 1 about 80 percent are above, in Fig. 2; and in Fig. 3 similar relations are found between colleges 2 and 20. Part of these differences may be due to curricular differences; but even if these differences were based on completely valid tests for all colleges, it would still not mean that the low-average college is necessarily inferior to the high in its *educational* contribution. Even though the achievement of its students *in physics* may be clearly inferior, the college may be giving these students other experiences more valuable to them, in view of their abilities, interests and social needs, than mastery of physics.

**College averages**

The fact that variability of scores within each college group is much more striking than differences between groups makes it necessary to interpret any distribution of college averages with caution. Recent reports[4] from the North Central Association of Colleges make it clear that college accreditation in the future will likely be along more constructive lines than in the past. College objectives and selection of students vary greatly. The North Central Association holds that these facts should be taken into account for accreditation. The same principle, we believe, should operate among physics departments. A physics department should not necessarily regard itself as superior to any other merely on the basis of a higher average score on the physics tests. The whole trend of the present study suggests that the best grounds for the relative ranking of departments might well be in terms of how well they diagnose the needs of students, how well they provide opportunity for the progress of their students at self-determined, differential rates and how adequately they define and attempt to meet their avowed objectives. Hours and units and credits and average scores furnish an excellent basis for self-study but are not remotely adequate for ranking the departments in order of merit for accreditation and like purposes. It is quite possible that some colleges of low average scores may be meeting the educational and cultural needs of their particular students more effectively than some colleges of high average score. The work of a college cannot be fairly judged except by taking into account the abilities and total educational needs of its own student groups.

Nevertheless, with this word of caution, distributions of college averages on the physics tests are presented in Table IV. Any college, by referring its average scores to this table, may identify its relative position among the participating groups.

**Gains in achievement**

The test results of those colleges administering tests before and after study were analyzed to discover the average gain of all students, without regard to type of school, and of student groups in colleges making the largest and smallest gains.

The amount of gain within single colleges varies markedly. These differences, however, are not subject to direct comparison, since the abilities, aptitudes and needs vary from college to college. The average gain for all students is of general interest; the gain of any one is of interest only to that college. It must be kept in mind that the average class gain must be interpreted in the light of local conditions. Each college must set up its own norms of expectancy and the deviations from this norm may be analyzed with reference to the individual case. Individual differences cover a wider range and are educationally more significant than the variations occurring among groups. Charting of single case gains can not be done in this study; it may be done, however, by each department to determine differences in growth. Doubtless some such charts will be startling. Within the same class, some students will make tremendous gains in achievement, others will make little or none. Those deviations which are most marked may be investigated with a view to further clinical diagnosis and educational guidance. But the average gain for the school is most significant when it is taken as the point of reference for the interpretation of the individual gain.

In this connection, it is difficult to exaggerate the importance of the variability of scores in the pre-study groups. Teachers in physics and

[4] North Central Association of Colleges and Secondary Schools, Commission on Institutions of Higher Education, *Statement of Policy*, February 3, 1934. M. E. Haggerty, *Accrediting Institutions of Higher Education*, The North Central Assoc. Quarterly, July, 1934. George F. Zook, *Accrediting Schools and Colleges*, The Educational Record, January, 1934.

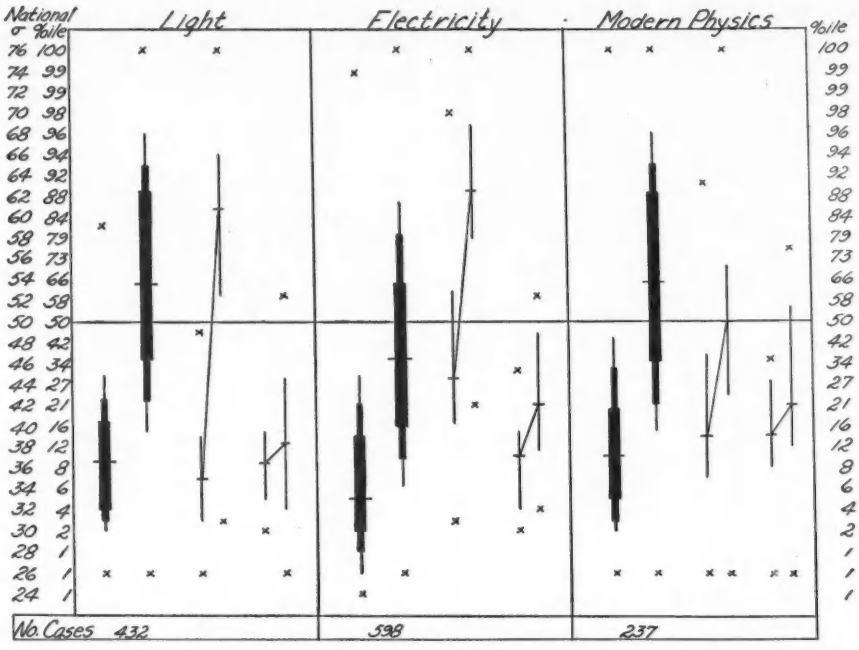


FIG. 4. Pre- and post-study comparisons for L, E and Mp. The two bars at the left of each section represent distributions of scores of all students who took the tests before and after study. The two following figures (light lines) represent the pre- and post-study results for the two colleges having the largest and the smallest gains in each topic. To be read in the same way as Fig. 2.

other subjects are only too painfully aware of the variability in post-study groups because there are too few classes in which some students are not "flunked;" but not many teachers are clearly aware of the highly important and encouraging fact that a few students at the *beginning* of a course are already able to equal or exceed the average of the whole class at the *end* of the course. Fig. 4 shows that in L, at least one student at the beginning was above the average of the class at the end of the semester; and in both E and Mp one student *at the beginning* secured a score which was achieved only by the highest 1 percent of the class *at the end*. In E about 5 percent of the pupils secured pre-study scores above the post-study average score of the entire group that took the tests.

The advantages of early identification of such promising pupils are obvious and it is gratifying to note that during the last few years the search for and special study of such pupils has been a growing concern of an increasing number of teachers. Of course, no single test can be expected to give a fully trustworthy index of even one of the many qualities required for a productive career in physics but any student who *before* formal instruction secures a higher score than the average student secures *after* formal instruction is at least worthy of being studied. Some of them very richly reward the teacher who vouchsafes them special attention and encouragement.

The students at the other end of the scale also merit special attention. Some of them after taking a full course, at the expense of the state or of their parents, are still in the score-range of the lowest 1 percent of pre-study norms. Some of these students work harder than the majority of those who make large gains and passing grades. Although it is obvious that they do not deserve, or rather that their lot will not be improved (and might be aggravated) by receiving passing grades *in physics*, it is the opinion of a growing number of educators that such students deserve something more constructive and helpful

TABLE V. *Numbers and mean scores of students reporting indicated professional goals.**

| Professional Goal Group | M | H | S | L | E | Mp | M+H | M+H+S | L+E | L+E+MP |
|---|---|---|---|---|---|---|---|---|---|---|
| Medicine | | | | | | | | | | |
| N | 1975 | 1941 | 1010 | 1202 | 1286 | 636 | 886 | 845 | 1113 | 612 |
| Mean | 17.0 | 10.7 | 7.8 | 12.3 | 17.2 | 6.1 | 26.9 | 35.6 | 29.4 | 36.7 |
| Law | | | | | | | | | | |
| N | 269 | 268 | 166 | 139 | 144 | 32 | 86 | 156 | 136 | 30 |
| Mean | 18.8 | 11.3 | 8.5 | 14.4 | 16.2 | 4.7 | 24.8 | 42.3 | 30.7 | 34.6 |
| Engineering | | | | | | | | | | |
| N | 2390 | 2039 | 978 | 1482 | 1529 | 370 | 1145 | 833 | 1220 | 357 |
| Mean | 19.9 | 11.7 | 8.1 | 12.7 | 18.4 | 6.8 | 32.1 | 39.6 | 31.2 | 39.6 |
| Architecture | | | | | | | | | | |
| N | 144 | 127 | 58 | 58 | 100 | 16 | 72 | 52 | 53 | 15 |
| Mean | 16.6 | 9.1 | 7.2 | 10.7 | 14.7 | 5.7 | 25.7 | 34.7 | 26.4 | 33.8 |
| Teaching | | | | | | | | | | |
| N | 902 | 886 | 380 | 597 | 551 | 225 | 486 | 313 | 510 | 178 |
| Mean | 17.7 | 10.3 | 8.4 | 12.0 | 17.0 | 6.5 | 26.3 | 39.3 | 28.8 | 38.0 |
| Ministry | | | | | | | | | | |
| N | 97 | 93 | 34 | 28 | 28 | 16 | 58 | 33 | 27 | 16 |
| Mean | 16.3 | 9.2 | 7.0 | 9.9 | 14.0 | 5.6 | 25.2 | 33.3 | 23.5 | 28.4 |
| Agriculture | | | | | | | | | | |
| N | 69 | 42 | 23 | 24 | 30 | 6 | 21 | 19 | 18 | 4 |
| Mean | 13.7 | 11.0 | 6.0 | 11.9 | 16.0 | 5.3 | 32.5 | 33.4 | 28.2 | 35.0 |
| Business | | | | | | | | | | |
| N | 514 | 492 | 307 | 302 | 301 | 102 | 183 | 287 | 262 | 92 |
| Mean | 16.8 | 9.6 | 7.3 | 12.3 | 15.2 | 4.4 | 24.5 | 35.1 | 28.1 | 30.5 |

* In view of the varying numbers of cases and the uncertainties of sampling, the indications of this table must be interpreted with great caution.

TABLE VI. *Numbers and mean scores of students in each college class.**

| College Class | M | H | S | L | E | Mp | M+H | M+H+S | L+E | L+E+Mp |
|---|---|---|---|---|---|---|---|---|---|---|
| Freshman | | | | | | | | | | |
| N | 1484 | 1290 | 772 | 735 | 834 | 263 | 483 | 723 | 706 | 241 |
| Mean | 17.2 | 10.3 | 7.9 | 11.9 | 16.4 | 5.6 | 26.3 | 37.5 | 28.8 | 34.0 |
| Sophomore | | | | | | | | | | |
| N | 4294 | 3975 | 1942 | 2725 | 2768 | 991 | 2070 | 1615 | 2316 | 929 |
| Mean | 18.9 | 11.3 | 8.2 | 12.6 | 18.2 | 6.3 | 29.9 | 38.6 | 30.5 | 38.8 |
| Junior | | | | | | | | | | |
| N | 1512 | 1518 | 771 | 1052 | 1052 | 393 | 671 | 687 | 912 | 366 |
| Mean | 18.0 | 11.8 | 8.0 | 13.2 | 16.6 | 6.6 | 27.3 | 38.5 | 30.0 | 37.1 |
| Senior | | | | | | | | | | |
| N | 486 | 487 | 198 | 277 | 348 | 136 | 258 | 158 | 243 | 125 |
| Mean | 16.6 | 10.3 | 8.0 | 11.2 | 14.8 | 5.9 | 26.0 | 36.0 | 26.4 | 33.3 |

* In view of the varying numbers of cases and the uncertainties of sampling, the indications of this table must be interpreted with great caution.

for their time and money than merely a failing grade. It is equally true that society deserves more for the time and money spent on such pupils than the mere stigmatizing of them as "failures." The question of what to do for such students can be answered only by careful and long continued experimentation, in which physicists will play their part. An essential part of such experimentation will be a surer identification of such students and a more adequate study of their growth along various lines of development. Growth cannot be studied without comparable measures; hence the emphasis which the committee has placed on the importance of developing *comparable* tests in physics.

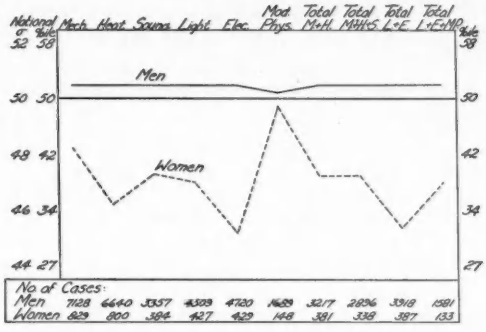


FIG. 5. Sex differences. Median scores for men and women are graphed in national percentiles scaled to sigma units. The numbers of cases are indicated at the bottom of the figure.

**Professional goal and college class groups**

Table V is interesting in that it reveals relatively slight differences in mean scores among the various professional goal groups. There are apparent trends, but the group differences are so slight as to be of doubtful practical significance. More startling still is the absence of mean differences among college classes, freshman to senior (Table VI). The drop in the senior year is quite similar to that found in surveys on subjects other than physics.

**Sex differences**

It is a well-known fact, derived from numerous studies, that certain characteristic differences in achievement occur between men and women. Women are generally superior in literature and language subjects but generally inferior in mathematics and science. The mean differences, however, seem to be small in relation to the variability of either sex group; the overlapping is usually quite extensive. Fig. 5 shows the medians of men and women on each topic of the physics test. It is difficult at this time to account surely for the fluctuations in medians for the women's group; but they are doubtless due in part to sampling differences. More than 800 women took the M test but only 148 took the Mp, which possibly accounts for the higher relative average of the women on this topic, since these 148 women may be much more highly selected than the larger group that took the other tests.

The data of Fig. 5 should not obscure the importance of individual scores. Within both sex groups individual differences are more important than group differences. The superior student, regardless of sex, should be identified, encouraged to continue in subjects for which he displays a particular flair and given special work if such seems desirable to the department head in the light of all relevant data and circumstances. The special encouragement and guidance of those few students who show a special gift for physics is considered a primary obligation and privilege by many teachers of physics. One department head has suggested that it may be a greater contribution to turn out one gifted research scholar than to turn out a hundred "pass" students each year. But there is no necessary conflict between these two types of contributions. Physics has rich cultural as well as disciplinary and scientific values; special care of the few gifted pupils does not mean that the less gifted pupils need to be neglected or sacrificed on the high standards which are appropriate only to the gifted ones.

## IV. THE QUESTIONARY RESULTS

**Attitude toward tests**

Of the participating departments, 120 answered in some detail the questionary "Criticism of Physics Tests." One item on this form called for a summary of opinion about the attitude of students and instructors toward the tests. As indicated by the completed blanks, the attitude was as follows: 92 departments reported "favor-

able," 7 "unfavorable," 8 "divided" and 14 failed to answer. Several questions were included in the blank for the expression of judgment on verbal formulation of items, length of the tests, etc. Table VII summarizes these opinions.

TABLE VII. *Summary of opinion on length of tests.**

| | M,H, S | L,E, Mp | M | H | S | L | E | Mp |
|---|---|---|---|---|---|---|---|---|
| Too long | 7 | 7 | 14 | 5 | 2 | 5 | 7 | |
| Too short | 5 | 5 | 4 | 7 | 7 | | 1 | 4 |
| Right length | 44 | 36 | 18 | 15 | 10 | 8 | 6 | 4 |

* The first two columns represent number of departments reporting undivided opinions on grouped first and second semester tests. The remaining columns represent divided opinion or opinions on less than the total number of subjects for either semester. Number of colleges reporting, 112.

There was very little agreement upon what questions in the entire battery were outside the realm of first year college physics. About 30 percent of the items were questioned by at least one commentator, but only those items on which more than one objection was raised will be indicated. In the following list the number of the item is indicated and after it, in parenthesis, the number of physicists who objected to it. The data are taken from 120 questionaries and refer to the 1933 forms of M, H and S and the 1934 forms of L, E and Mp:

M 30 (3); H 17 (2), H 19 (4), H 20 (4), H 22 (2), H 24 (3), H 27 (5); S 14 (4); L 26 (2), L 27 (4), L 32 (3), L 33 (3), L 34 (2), L 35 (4); E 24 (4), E 39 (4), E 40 (2), E 42 (5), E 44 (2); Mp 10 (2), Mp 12 (3), Mp 14 (2).

For comparison of these judgments with the experimental findings, see Table I.

This report is, of course, not the place to present an extended account of the detailed comment and criticism relative to specific items in the tests. But the committee wishes to acknowledge its indebtedness to those members of departments who gave their opinions on individual items and suggested ways and means for improving the content of the tests and the method of expression.

**Cross section of opinion on program**

Because many opinions are interesting in their own right, a brief cross section of comment classed under four headings has been attempted below. These comments were selected by the usual process of sampling, from more than 1000 letters. Since the remarks made are strictly confidential, the names of the writers are withheld from the specific quotations; but these names are listed alphabetically for the sake of general identification. In a few instances the quotations have been slightly modified for simplicity; and all, of course, are apart from their original contexts.

Vola P. Barton, Goucher; Frederick L. Brown, University of Virginia; O. H. Blackwood, University of Pittsburgh; W. G. Cady, Wesleyan University; L. A. DuBridge, Washington University; John R. Hobbie, Skidmore; H. V. Houseman, Menlo Junior College; L. R. Ingersoll, Wisconsin; A. T. Jones, Smith; E. R. Laird, Mount Holyoke; Harvey B. Lemon, Chicago; T. J. Love, Loyola College; Raymond McElligott, U. S. Coast Guard Academy; F. J. Mellencamp, Wisconsin State Teachers College; Frederic Palmer, Jr., Haverford; R. L. Petry, University of the South; Grant L. Pistorius, St. Joseph Junior College; R. A. Porter, Syracuse; M. H. Trytten, Pittsburgh, Johnstown Branch; J. A. Van den Akker, Washington University; Calvin N. Warfield, North Carolina; E. H. Warner, Arizona; Dorothy W. Weeks, Wilson College; S. R. Williams, Amherst; L. Wilson and D. Heyworth, Wellesley; and Jay W. Woodrow, Iowa State College.

(a) *General comment.* "We have appreciated the opportunity of using these tests. There is no doubt in my mind that they are extremely useful in enabling us to check up on our students as compared with those in other institutions."—"I am quite pleased with the questions. The committee that formulated them did a fine piece of work."—"The tests are interesting and most helpful. The results should be of value to those who are teaching in small colleges."—"The danger of these tests is paradoxically that they become too successful. They may become a standard which will influence teaching to such an extent that instructors will bend their efforts to making higher and higher scores."—"We have been quite favorably impressed with these tests. We plan to use them again another year for all three quarters."—"I wish the committee continued success in carrying on this important work."—"These examinations are very well prepared and stimulating and I trust that next year another set will be available."—"The tests are very good as a whole, far superior to any objective tests commercially available."—"I consider these the best set of College Physics tests I have ever seen."

(b) *Student reactions.* "Students like taking the tests for the most part but had no comprehension of how poorly they answered them."—"Few if any complaints from students. They seem to prefer the tests to the traditional examinations and instructors are well satisfied."—"The reaction of the students to the tests was splendid. They showed keen interest, enjoyed the exercise and were well satisfied with the results."—"Students seem about equally divided; those in favor, however, are enthusiastic, while those against this type of examination are not strongly against it and usually have no tangible objections to offer."—"On the whole we find the tests conducive to quick, accurate thinking."—

"The students have not been antagonistic but think that the tests are long and hence do not really show what one can do in a thoughtful way."—"Students felt that this part of our examination was a considerable strain. They hoped that we should not use an 'imported' test at the end of the second semester."—"The students were very much interested in tests which would allow their work to be compared with that done elsewhere."—"When we can get our students acquainted with the nature of these tests and have them take the right attitude, we shall be glad to participate more fully."

(c) *Difficulty of the tests and time limits.* "The tests were too short and too easy since 6 out of 37 men answered all of the questions on both tests, 2 more all of those on M and 4 more all of those on H. The 6 men who answered all the questions obtained raw scores of 72, 69, 68, 66, 55, 50. The man who scored 72 required only 50 instead of 90 min. In the future such tests should have at least 100 questions with 5 options to each question and the 25 additional questions should be of the more difficult type."—"The time allowance for the tests on L and Mp seems approximately correct. However, the time for the test on E might well be increased to 60 min. This would result in a better spread of grades and still leave a fairly wide margin at the top for only very superior students."—"Very few students can make scores above 60 on E and Mp. Why not ease up the tests a bit so that the better students will make 85, thus spreading out the grades more for better differentiation?"—"My first impression was that the tests were too long but national percentiles indicate otherwise."—"The tests may put somewhat too much of a premium on speed at this length. However, I think that they form a very valuable supplement to the usual examinations."—"The Mp test was too difficult."—"The examinations cover too extensive a field for the general course. I would prefer one with questions under three divisions: A, for general students in a course for information and survey only; B, for students in a college course intended to give a thorough grounding in the fundamentals of physics; C, for technical or engineering students."—"Where the plan was thoroughly explained and specimen questions were shown some days before the test, there was a 'chorus' of approval shown in unsigned essays. Where explanation was given on the day of the tests, two-thirds of the essays showed disapproval. The principal objections were that matter was presented which is not given in the text; and that the tests required so much close reasoning as to be confusing to nervous students. In general, I approve of the tests but believe that they would be improved by decreasing the amount of reading."

(d) *Use of test results.* "My question, I suppose, boils down to the old trouble of having to give marks and degrees whereas if we could but free ourselves from them and devote our attention to helping students learn and to arousing their interest it would be better for them and pleasanter for us. You are on the right track when you recommend separation of records—achievement, personality and work interests. If you can get this idea started, and I am at least one example to prove that you can, I hope you will go ahead with it."—"I do not favor such tests as exclusive means of determining grades. Students should be expected to answer questions requiring more sustained and extensive thinking."—"Mark answer correct only when *result and reasoning* are correct. We have tried this on several occasions, and have found that correct answer and correct reasoning very often do not go together."—"I am thoroughly in accord with the physics project but beyond the selective value of such examinations there is just one way to test a man—put him on some project as an individual and let some one with a flair for teaching supervise."—"In my course I emphasize problem solution. All quizzes are problem quizzes with open books. Your examinations have shown me that such a procedure does not sufficiently emphasize fundamental, theoretical items. I am ashamed to have my class rank low compared with other universities but I think this emphasis on problems is partially responsible for this fact."

## Differences in college averages related to various local conditions and factors

At the request of many participating colleges, evidence has been sought on the possible influence on college average score of such factors as (a) prerequisites, (b) time requirements in lecture and laboratory, (c) class size, (d) textbook used, and (e) type of institution. The only data available to the committee for the study of these complex questions were the materials sent in by 116 of the 355 colleges in the form of a "Report on Course in Physics." Although it is obvious that the committee has no control whatever over the size or representativeness of the samplings of colleges that fall in one or another group with respect to any of the factors named above, it was thought worth while, as a first step, and in view of the expressed interest of many colleges, to compare the averages of the colleges in each such group. If any significant differences had been found, they might be due to one or more of a number of other factors not included in the available data, and therefore beyond the statistical control of the committee. As a matter of fact the differences found are so small and inconsistent as to be almost negligible, and certainly inconclusive. If the factors studied do have significant influences, these influences are effectively overridden and concealed by other influences not within the statistical control of the committee.

(a) *Prerequisites.* The largest and least inconsistent differences were found between the groups of colleges having different prerequisites for admission to the physics

course. Considering the average on all six physics tests, the college groups ranked approximately as follows:

| Group | Prerequisites | Average position relative to national average |
|---|---|---|
| 1 | Alg., trig. and H. S. physics | About .2$\sigma$ above |
| 2 | Alg. and trig. | About .1$\sigma$ above |
| 3 | No prerequisite | About .1$\sigma$ below |
| 4 | Algebra | About .2$\sigma$ below |

If we rule out sampling and all the other factors that might easily account for these differences, then we may tentatively conclude that the requirement of algebra, trigonometry and high school physics favors higher averages in elementary college physics but that requiring algebra tends toward lower averages than does no prerequisite for admission to the college physics course. The first indication seems to be as reasonable as the second is questionable.

(b) *Time requirements in lecture and laboratory.* The differences are so slight and inconsistent that no indications can be detected.

(c) *Textbook used.* The differences are negligible and inconsistent.

(d) *Types of institutions.* Two sets of comparisons were made. In the first, the colleges were grouped according to the professional and educational plans of the majority of their students—pre-engineering, premedical, teaching and general. The differences are small and irregular but the pre-engineering averages are generally highest and the general and teacher groups lowest. In the second, public, private, men's, women's, coeducational, denominational and non-denominational college groups were compared. The differences are entirely negligible, except that women's colleges are lowest.

(e) *Class size.* The correlations between class size and average scores on the M and E tests ranged from $-.23$ to $+.11$ with large probable errors, the numbers of cases (colleges) ranging from 54 to 84.

On the whole it appears that none of the factors studied here, with the possible exception of prerequisites, has any noticeable influence on the average scores of the colleges. If any of these factors do have significant influence, that influence is effectively concealed by other and stronger factors. Nevertheless, we know, from Table IV, that the average scores of colleges do vary markedly. In M, for example, the percentile score of one college is as high as 97, while that of another is as low as 8. If these differences are not accounted for by such external conditions as size of class, text used, time requirements in lectures and laboratory, type of institution or prerequisites, to a notable degree, it is more than probable that they are largely conditioned by the original selection of students. Differences in efficiency of instruction cannot be clearly invoked because it is obvious that some students in the lowest-average colleges achieve high scores and many students in the highest-average colleges achieve low scores. The general conclusion suggested by these data, and by other considerations, is that the crux of achievement in physics, as in other realms, depends to a significant degree upon the intelligence, aptitude, interest and initiative of the individual student—in short, upon all those factors of individuality which can be discovered and studied most effectively by means of the clinical method. In a word, individual rather than group diagnosis appears to be the more promising lead for the purposes of this study of the learning and teaching of physics in American colleges.

Although the committee frankly admits its present inability to isolate factors of course organization which may have contributed to differences in average class performance, it nevertheless sympathizes strongly with the departments in their desire to uncover whatever relationships may exist. To make a thoroughgoing experimental attack upon the isolation of such variables would necessitate much more information than is now available. For such a study it would be necessary to have scores on some standard intelligence test, relative rank in high school class for each student tested and carefully prepared answers to all the questions which appeared in the form "Report on Course in Physics." Other types of information would likewise be necessary but enough has been mentioned to illustrate what a complex job of reporting would fall upon participating departments if the attempt to isolate the variables contributing to average class differences were undertaken in detail.

Such a study, of course, is not impossible; but the simplest and most direct attack upon it is one which will not burden departments with elaborate reporting of information. This plan demands only that the majority of departments (a) give a scholastic aptitude test, such as the Thurstone Psychological Test, to all students near the beginning of the elementary physics course and (b) give the physics tests twice each

semester, once before the course is taken and once after it is completed. If the majority of departments will do this pre-testing, the average results for each department would likely be the best possible single criterion of selection available. And class averages on the pre-tests compared with similar averages on post-tests would yield a reliable measure of departmental gains over the period in question. Those departments making very unusual gains might then describe their course organization in the report and suggest what in their opinions are the most significant contributing factors. If special departmental studies of this type can be included in the report for the coming year, such information together with the relative gains made may be of considerable value in shedding light on the very complex problem involved in the relation between course organization and average departmental scores. Other suggestions for a study of the problem described will be gladly entertained.

In view of the tremendous range of individual differences in student ability, it is unwise to measure the competency of the instructor by the class mean score, which depends so much on the selection of pupils. A far better index of instructional effectiveness is the manner in which individual cases are handled, the accuracy with which successful workers are selected and the consideration shown to failing students. No matter how brilliant the instructor, it is almost impossible for him to raise materially the average achievement of a large, *unselected* class; however, he can identify students who give promise of rapid improvement and he can direct many others into fruitful and interesting lines of endeavor. It should be exceedingly clear that what is recommended is *not a lessening of teacher effort* but a change in objectives. The concept of "the great teacher" should not be abandoned; rather it should be revitalized and reinterpreted.

## Reply to minority criticism

(1) *Length of tests.* Obviously, if only one question of average difficulty is put to a group of students and it is scored right or wrong, we should expect 50 percent to achieve a perfect score and 50 percent to receive a zero score. However, we know that the distribution of the students' ability in the course work from which the question was drawn does not fall into two discrete groups. Very few cases, even in groups of thousands, have ability equal to zero or to 100. Actually, the greatest frequency will occur at or near the average, the frequencies decreasing gradually on both sides of this point, so that if the frequencies are plotted a figure approximating the Gaussian curve is described. In order adequately to measure his students' achievement, then, the instructor must devise a test which will yield a wide range of scores. Only in this way can the relative ability of any given student be decided. An examination consisting of 1 question only does not serve this purpose, since the student who has mastered 90 percent of the course material still has a 10 percent probability of failing. If the number of questions is increased to 10, this probability is decreased to 1 percent and if the number is still further increased to 100, 96 out of 100 of the very well-informed students will achieve scores between 85 and 100 and the scores of the remaining 4 will fall between 75 and 84. Thus, not only are the superior students reliably differentiated from the mediocre students but gradations of ability in both groups may be determined.[5]

(2) *Time limits.* In reference to the time allowance for each test, it has been experimentally verified that, beyond a certain point, an increase in time allowed does not result in a significant rank difference in results. Only in very rare cases, when the student is a pathologically slow reader, for instance, is the score or rank materially altered.

(3) *Difficulty.* The presence of items in the tests which are beyond the ability of the majority of the students, is justified by the fact that only by means of such critical questions can fine distinctions be made between very good students. If the tests were such that the average student could achieve a high score, the markedly superior student would not be adequately differentiated; his score would not be a reliable index of his ability; and, in consequence, it would be of less use in the diagnosing or advising work of the department. The early identification of brilliant students is of fundamental importance in the later selection of research workers or candidates for advanced degrees and in the recommendation of graduates to positions in industry and business. Test scores alone are not reliable for such selection; but taken in conjunction with other types of information, they are significant and they are the only stable common denominator available.

(4) *Test fractionation.* Several departments have requested that separate tests be compiled for engineering, medical and general students. The advantages deriving from such distinction, however, could be secured only by large expenditures for which the committee does not have the resources. If any differentiation were found to be advisable, separate norms for each professional goal group might be established on the same test. But because the differences which were found to exist among such groups were so slight as to be of no practical significance, any differentiation of examinations appears to be unnecessary. Table V presents the data upon which these conclusions are based.

---

[5] For a more complete discussion of this problem, see C. Posey, *Luck and Examination Grades*, J. Eng. Ed. **23**, 292 (1932). Reprints may be secured on request. For an abstract of this article, see Am. Phys. Teacher **1**, 31 (1933).

(5) *Test uses.* The committee wishes to make its position clear in reference to the use of objective tests. Their exclusive use is in no case recommended but as a supplement to the essay-type examination they are invaluable. It is true that in objective testing, no opportunity is afforded for verbal expression or a written organization of ideas. On the other hand, they are the more competent measures of precise knowledge and of ability to apply such knowledge to problematical situations. Moreover, the ability to verbalize ideas and pure factual knowledge have been found to correlate highly. And certainly the objective test can be more exactly scored because it is less sensitive to the subjective judgments incident to scoring the essay-type examination. The ideal combination, of course, is the use of both types of examinations, each as a supplement to the other, the scores on each being recorded separately.

## V. Conclusions

This opportunity is taken to acknowledge an obligation to the cooperating physics departments for the time which they have spent and the interest which they have displayed in making the program a success. Thanks are due especially to those instructors who have offered criticism and suggestions toward the improvement of future physics tests.

The committee calls attention to the fact that drawing conclusions from a report of this sort is a very difficult undertaking. The variables involved are much more complex and much less amenable to quantitative analysis, than those occurring in purely physical research. Accordingly, the generalizations which follow are to be regarded in the light of these complexities, and must be taken more as suggestions than as scientific conclusions. With these cautions definitely in mind, we may proceed to a brief summary of the points presented above.

By means of the college averages each school may compare its achievement with that of any other. However, this was not the major purpose of the physics program. It was set up primarily to determine those variables which would be diagnostic—both for the instructor and the student. To this end variability in achievement was measured; national percentile tables were computed to enable ranking the individual's score on national norms; comparable forms of the same test were used to permit a study of gains in achievement; and group differences were charted and analyzed. It has been noted that colleges differ markedly in average achievement and in average gains in achievement as measured by pre- and post-study testing. But it has been urged that interpretation of such results be made in the light of local conditions, that individual deviations from the class average are even more important than college deviations from the national averages and that group gains are of most value taken as local norms of expectancy. Group differences are conspicuous by their absence, the variation between men and women being the only one of sufficient magnitude to be remarked. Even the variation in mean scores of different professional goal groups is practically insignificant. Nor does group organization, in any respect, influence to a marked degree class achievement. By means of the results derived from this study, the importance of individual differences in student capacity has been demonstrated and the need for individual diagnosis emphasized.

This brief summary of generalizations which has been drawn from the tabular data is, in a sense, rather negative in its implications. It tends to call into question the reliance which has heretofore been placed upon prerequisites and general classroom and laboratory procedures designed for *group* instruction. Although the data reported do not point to any one specific change in attitude toward teaching, nevertheless they suggest that some reorientation is in order.

Teaching, like the practice of medicine, is still an art. But the various forms of measurement at present available both to teachers and to physicians give these two rather similar disciplines opportunity to make use of scientific method. Just as the physician employs chemical analysis, the x-ray, blood count, metabolism tests and the like, so the teacher makes use of test results, personality ratings and anecdotal accounts of academic and work history. When the physician or the teacher finally prescribes for his case he makes use of the total clinical picture rather than the results of any single interview or any single test. And so far as he makes use of the total picture he is an artist rather than a scientist; but he is scientific to the extent that he uses facts and sound methods for their interpretation.

No simple device for measuring any given variable will solve the problem of instruction.

But, on the other hand, the mere exercise of opinion without reference to facts would identify the teacher with the medical practitioner of many years ago. For the teacher the first significant step in a scientific direction is probably that of breaking down his composite grade into at least three elements. This involves recording: first, the achievement of students in terms of quantitative measures of their knowledge of and ability to think in terms of the subject matter; second, their peculiar character traits, their degree of initiative or, in rare instances, their "infinite capacity for taking pains" or, again, their tendencies to carry on independent, self-initiated study; third, their socio-economic history, at least for those students who give any promise of unusual understanding of the subject. Such records kept over a period of time should prove vastly significant for the selection of students for advanced or graduate work and for placing graduates in responsible positions in industry, in teaching or in research.

With thanks for the excellent spirit of experimental cooperation, the committee submits this report to the departments collaborating in the research. Obviously numerous problems of education remain unsolved; in fact, the significant feature of this report is that it has clarified some problems rather than that it has solved any. Continued experimentation is in point. It is to be hoped that even larger numbers will take part in continuing this program next year and that meanwhile departments will make use of the results reported for individual diagnosis and guidance. Before the report is made for the coming year, each department participating in the present survey is requested to inform the committee concerning any findings or suggestions which may be of general interest along the lines of individual diagnosis and guidance.

THE COMMITTEE ON TESTS OF THE
AMERICAN ASSOCIATION OF
PHYSICS TEACHERS.
*H. W. Farwell,*
*C. J. Lapp, Chairman,*
*Frederic Palmer, Jr.,*
*John T. Tate,*
*A. G. Worthing.*

*Additional copies of this Supplement may be obtained from the Committee on Educational Testing, F. S. Beers, Secretary, University of Minnesota, Minneapolis, Minnesota.*

*The American Physics Teacher* is published quarterly for the American Association of Physics Teachers by the American Institute of Physics, 11 East 38th Street, New York, New York.